

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1999 - 11th Annual Conference Proceedings

STATISTICS IN THE NEW MILLENNIUM: SOME PERSONAL VIEWS

Bryan F.J. Manly

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Manly, Bryan F.J. (1999). "STATISTICS IN THE NEW MILLENNIUM: SOME PERSONAL VIEWS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1256>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

STATISTICS IN THE NEW MILLENNIUM: SOME PERSONAL VIEWS**Bryan F.J. Manly****University of Otago, P.O. Box 56, Dunedin, New Zealand****Abstract**

In this paper I discuss three topics that I believe are relevant to the subject of statistics in the new millennium: (a) the impact of computers, and the state of computer-intensive methods as far as practical applications of statistics are concerned; (b) methods for the analysis of the extremely large data sets that are now becoming available; and (c) the use of statistics by scientists in general. For the first topic I suggest that the main advantage of computer-intensive methods is that they can under certain circumstances give simple and believable answers to questions when other methods fail. However, I caution against the uncritical use of computer power without proper checks that analyses work, particularly when conclusions are dependent on very complicated models with many assumptions that are difficult or impossible to verify. For the second topic I note that statistics grew up as a means of extracting the maximum amount of information from small sets of data, and we are now having some difficulty in adapting methods to huge data sets because sometimes the analyses that we might want to do are not possible even with today's powerful computers. I discuss this particularly in terms of the analysis of resource selection data by animals where geographical information system data are available to describe what is available for animals to use. For the third topic I suggest that statistics and statisticians have something of an 'image' problem with scientists in general. Many scientists do not appear to regard statistics as important for their discipline, and yet errors in the analysis and interpretation of data seem to be fairly common in the scientific literature.

Keywords: Bootstrapping; Computer-intensive statistics; Extremely large data sets; Geographical Information Systems; Randomization test

1. Introduction

This paper gives some personal views about the state of statistics as we approach and move into the new millennium (leaving aside the question of whether this begins in 2000 or 2001). I focus on three issues that I think are important in this respect.

The first issue concerns the impact of computers, and the use of computer-intensive methods. My thesis here is that the availability of considerable computer power, and the methods of statistical analysis that this permits, have been of tremendous benefit to statisticians. The tools of randomization, bootstrapping and Monte Carlo simulation are becoming part of the standard toolbox for the practising statistician, enabling solutions to be obtained for many problems that are very difficult to solve using traditional methods. However, I caution that there does seem to be a tendency

to use these methods uncritically. Also, I view the main advantage of these methods as being their potential to provide relatively simple solutions to complex problems. Unfortunately, they sometimes seem to be used to provide complex solutions to relatively simple problems!

The second issue concerns the development of methods for analysing the extremely large data sets that are becoming increasingly available. To some extent the problem here seems to be that the development of statistical methods during most of the twentieth century has been in the context of extracting as much information as possible from rather small data sets. Now some data sets are so huge that subsamples of the data have to be taken to get a reasonable number of observations for an analysis. Research into methods for extracting information from very large data sets is being carried out. However, my concern is that it appears that most of this research is being done by computer and information scientists rather than statisticians. This leads me to wonder whether the statistics profession in general is too constrained by its small sample history.

Finally, the third issue that I address is the image of statistics as far as the general public is concerned, and, more particularly, the image of statistics held by the scientific community. I give some examples which suggest to me that these images are not as good as they should be. Many scientists appear to act as if statistics and statisticians are largely irrelevant to their subject, and yet mistakes in the analysis and interpretation of scientific data seem to occur frequently. If my interpretation of the current situation is correct, then there is no easy fix. But this is a problem that needs attention because unless statistics is seen as being useful and relevant by the scientific community (who should be our best friends), then the long-term future for the subject is bleak.

2. The Impact of Computers and the State of Computer-Intensive Methods

The tools of randomization, bootstrapping and Monte Carlo simulation allow statisticians, and scientists in general, to solve some problems that are difficult to solve any other way. However, like other powerful tools, they need to be used with caution. A few examples will clarify what I mean by this.

To start with, consider randomization tests. These have a relatively long history, starting with R.A. Fisher's tests for a paired comparison design and the comparison of two sample means (Fisher, 1935, 1936). The basic idea is very simple. Thus, for the comparison of the mean values of two samples with sizes n_1 and n_2 , the observed mean difference is compared with the distribution of the mean difference that is generated by taking the $n_1 + n_2$ observations and randomly allocating them to the two samples, keeping the sample sizes fixed at n_1 and n_2 . The null hypothesis is that the mechanism generating the sample data is effectively equivalent to a random allocation. Therefore, if the observed mean difference is unlikely to occur from a random allocation then this is evidence that the null hypothesis is not true.

With 'pure' randomization tests like the ones proposed by Fisher there is nothing to go wrong. The size (percentage of significant results when the null hypothesis is true) of a test carried out with

a significance level of $100\alpha\%$ is exactly $100\alpha\%$, even when the observed mean difference is only compared with a small number of mean differences from randomized data (Dwass, 1957). Unfortunately, however, these 'pure' tests have only very limited applications in practice so that it becomes desirable to extend the randomization argument to more complex situations. This tends to lead to unforeseen complications, and possibly to tests that do not have the expected properties.

A case in point is Raup and Sepkoski's (1984) randomization test for periodicity in fossil extinction rates over the last 250 million years. This test appeared to give strong evidence of a 26 million year cycle in extinction rates, leading to many attempts to find an explanation for such a long-term cycle. However, subsequent analyses that are summarised by Manly (1997, Section 11.5) have led to serious doubts about the validity of the test because of a time trend in the magnitude of extinction rates, and the lack of an allowance for the multiple testing involved in considering possible trends from 12 to 60 million years in length. This example demonstrates how a seemingly simple and straightforward randomization test may actually have hidden flaws.

Bootstrapping is currently more popular than randomization as a method for analysing data. The original developments of bootstrap theory had a heavy emphasis on methods for producing confidence limits for unknown parameters. Bootstrap tests of hypotheses can also be made, but these have not always been set up correctly, even with papers published in statistics journals (Hall and Wilson, 1991).

My concern about bootstrapping as currently used is that too often the checks for whether it actually works are minimal, or non-existent. For example, many papers in statistics journals end up with some complicated estimator, and then just say something like 'variances were obtained by bootstrapping'. The fact is that bootstrapping may not actually work very well, as can be demonstrated by trying to obtain valid confidence intervals for the mean and standard deviation of an exponential distribution from a random sample of size 20 (Manly, 1997, Chapter 3). To me, using bootstrapping in a new situation without testing it out properly seems like a chemist devising a new chemical analysis without making sure that it measures what it is supposed to measure.

Monte Carlo simulation covers such a wide range of different techniques that it is difficult to comment on this in a few words. All I will say here is that I am rather suspicious of the results from the currently popular Markov Chain Monte Carlo Bayesian analyses where there is an extremely complicated model, many assumptions, and not much data.

I will end this section by giving two examples of computer-intensive methods. I regard the first as a 'nice' example because it uses computer power to produce a simple answer to an important question, with a fairly complicated set of data. The analysis follows the KISS philosophy (Keep It Simple Statistician). The second example is of an analysis that I find unconvincing. This follows what I call the BOYS philosophy (Bayes Overcomes Your Shortcomings). The problem here are that there is very little data and the analysis depends on a very large number of unverifiable assumptions.

Example 1: The Effect of 1080 Poison Pellets on Invertebrates

Poison pellets using sodium monofluoroacetate (1080) are laid in New Zealand forests to control possum numbers, because these animals are a major pest killing native forests. Unfortunately, there is some concern that these pellets may be killing invertebrates as well. To examine this question, the government Department of Conservation has run some experiments.

Trials were run for either 9 or 18 days, using neighboring grids of pellets, with one grid as a control receiving placebo pellets, and the other grid receiving 1080 poison pellets. Pellets were replaced every day, and the data obtained were counts of the live invertebrates found on the pellets at a certain time each day. For the 9 day trials pellets were put out as shown in Table 1 (a). For the 18 day trials the same design was used but was repeated for another 9 days, as shown in Table 1 (b). More details about the experimental arrangements are given by Sherley and Wakelin (1999).

There are many different approaches that might be used to analyse the data from this experiment including, for example, attempting to model the counts of invertebrates on individual pellets. However, the method that was adopted was a fairly straightforward randomization test for the effect of 1080. It is important to realize in this respect that this was an initial experiment to determine whether there is any effect at all. Therefore, estimating the magnitude of an effect to a specified accuracy was not a major consideration.

To carry out a randomization test it is first necessary to measure the effect of poisoning, if any. A simple way to do this is illustrated in Table 2, based on the results for days 1 to 9 of the first trial. Two measures are calculated. The first is E_1 , which is the change in the treated - control difference between the initial 3 days of the experiment (when no poison was laid) and the next 3 days (when 1080 pellets were laid on the treated grid). The treated - control difference was initially 0.35, but changed to -0.39, to give $E_1 = -0.39 - 0.35 = -0.74$. Thus the apparent effect of the 1080 pellets was to reduce the invertebrate count by an average of 0.74 per pellet. This is a measure of the immediate effect of using 1080 pellets. Presumably, the reduction means that the invertebrates were either poisoned and died elsewhere, or they were repelled by the poison.

The second measure is E_2 , which is the change in the treated - control difference between the initial 3 days and day 9. The initial difference of 0.35 changed to -1.13, giving $E_2 = -1.13 - 0.35 = -1.47$. This is a measure of the residual effect of the treatment after it has been discontinued for 3 days.

When the trial was repeated on days 10 to 18 as shown in Table 1 (b) it is possible to calculate another two statistics E_3 and E_4 . Here E_3 is the same as E_1 but calculated for the second 9 day period instead of the first 9 day period. It measures the effect of the treatment when it is repeated. Similarly, E_4 is the same as E_2 but calculated for the second 9 day period. This measures the residual effect for a second treatment.

Table 3 shows the full experimental results for E_1 to E_4 for the 13 trials carried out. The mean values of these statistics are shown at the foot of the table, and the question to be considered is whether these mean values indicate a significant impact of the poison pellets.

To run a randomization test it is only necessary to note that switching round the control and treated data for one trial just has the effect of changing the signs for each of the statistics E_1 to E_4 . It can therefore be argued that on the null hypothesis of no treatment effect the signs in any row of Table 3 were equally likely to be as they are shown, or reversed. The test therefore involves seeing whether the mean values shown in the table are significantly low in comparison with the distributions found by randomly reversing the signs for E_1 to E_4 for the individual trials with probability 0.5. A one sided-test is called for because it is hard to imagine how the use of 1080 pellets could increase invertebrate numbers.

The @RISK software (Pallisade, 1995) was used to generate 10,000 randomized sets of data. The proportion (p) of times that a mean as low as that shown in Table 3 was then determined, for each of the mean values for E_1 to E_4 . The probability of observing a mean as low as E_1 was very small ($p = 0.0001$), giving very strong evidence of an immediate reduction in invertebrates from the poison pellets. The mean of E_2 was not significantly less than zero ($p = 0.19$), so there is no real evidence of a residual effect from one treatment. The mean of E_3 was significantly less than zero ($p = 0.047$), giving some evidence of a reduction in invertebrate numbers following a second treatment. The mean of E_4 was quite significantly less than zero ($p = 0.015$), giving some evidence of a residual effect of the second treatment.

This example is really a complicated application of Fisher's (1935) randomization analysis for the method of paired comparisons. There is more that could be said about it, plus some questions such as why the experimental design was chosen. It seems a bit ad-hoc. Also, of course, many statisticians would prefer a model-based analysis using the raw data rather than the summary statistics E_1 to E_4 . I like the analysis given here because it is easy to understand, and quite convincing. Some other examples of a similar nature are described by Clarke *et al.* (1998), Cushman *et al.* (1993) and Hayes (1995).

Example 2: The Effect of Fisheries Bycatch on New Zealand Sea Lions

With my second example it is not possible to do more than briefly sketch the situation. The question of interest to the New Zealand Department of Conservation is the effect of fisheries bycatch on the New Zealand sea lion population, which mainly breeds on the Auckland Islands to the south of the South Island of New Zealand. In the past few years squid fishing has resulted in about 50 sea lions being killed each year in the fishing nets. Government regulations restrict the number of deaths that are allowed to occur, with the fishing grounds being closed down when the allowed number is exceeded. This has led to a good deal of controversy over how the maximum number should be fixed.

The only real data that are available for an analysis are counts of sea lion pups on beaches that have been carried out at irregular times since 1942. For the analysis being considered here only 51 counts were available, of which only 7 recent ones are considered to be reliable because the counting methods used are well documented. There is little knowledge about the sea lion population in terms of survival rates, reproduction rates, dispersal rates between breeding colonies, or the population size before humans arrived in the area.

For this complex situation an analysis was recently conducted by a leading team of fisheries scientists (Hilborn et al., 1997). Briefly, the model involved a likelihood constructed by assuming normal distributions for pup counts, with guessed standard deviations, unverifiable models for density-dependent reproduction and dispersal, and there were 22 parameters, with 20 independent prior distributions and 2 fixed values. Posterior probability distributions for parameters, and the distributions for future values of parameters were determined by Markov Chain Monte Carlo, and the authors claimed to produce probabilities for new colonies being formed, population extinction, etc., for different bycatch control strategies.

My impression is that analyses of this type are becoming very popular. My opinion is that the outputs are not believable unless there is a very substantial study to see how the results are altered when assumptions are changed. I am reminded of what George Bernard Shaw (1944) had to say about certain biometricians, many years ago:

Through my acquaintance with the late Karl Pearson I was early impressed with the necessity of mathematics in biology and indeed in all branches of science. Pearson, always smiling and charming, would not admit that anyone who was not a mathematician could claim any scientific authority whatsoever. I subscribed faithfully to his journal Biometrika without understanding any of its equations or more than, say, five percent of its sentences.

But I found that the biometricians, though their technical skill and subtlety seemed wonderful to me, were as credulous, as prejudiced as thoughtless as to the facts they were measuring and the assumptions from which they started, as Isaac Newton himself. Even their counting was not to be depended upon; for they added up facts and opinions indiscriminately, and cooked their calculations by 'weighting' them with fancy figures which represented nothing but their personal guesses and tastes.

The reference to Isaac Newton has to do with the fact that at one time Newton spent a great deal of time in developing a chronology of the world from its beginning in 4004 BC. Many of the opinions expressed in Shaw's (1944) book have proved with time to be quite wrong. However, one has to admit that he had a good way of expressing things.

3. Extremely Large Data Sets

Recently I received an e-mail from the British statistician John Jeffers. Referring to the consulting scene in the UK, he had this to say:

Despite some 40 years trying to teach statistics and modelling to scientists and environmental professionals, the same old mistakes are being made, but they are worse because it is now possible to collect massive amounts of data before anybody realises that they don't know what to do with them.

This is a good introduction to the second topic that I wish to discuss.

My contention is that the early days of statistics were characterised by an interest in getting the maximum amount of data from small data sets. However, now the problem often is that data sets are so large that statistical analyses have to be done on subsamples of the data. To illustrate what I mean, I will consider another example.

Example 3: Moose Winter Habitat Selection on the Innoko Wildlife Refuge in 1994 and 1996.

This example concerns moose winter habitat selection in 1994 and 1996 in the Innoko River Corridor Stratum, an area of about 500 square miles. Habitat information came from 30m by 30m pixels in a geographical information system (GIS), but initially moose were assumed to select circles with a 400m radius, consisting of 561 pixels, for their use. Thus the habitat available to the animals was assumed to consist of all such circles. Used circles were assumed to be centered where moose were found. Every pixel was classified into 22 land cover classes based on vegetation characteristics, so that each 400m circle can be described by the covariates X_1 to X_{22} , where X_i is the proportion of habitat i . A fuller description of the study details is given by Erickson *et al.* (1998).

A simple analysis of the data involves considering every possible pixel as the center of a circle that could be used by the moose. Logistic regression could then be used to estimate the probability of a moose group using a circle with particular values for the covariates. The problem with this is that there are over 1 million available circles, of which 67 are used. With such a large data set logistic regression does not seem to be possible.

An alternative approach involves comparing the used circles with a random sample of available circles. It seems a pity not to use all the available data, but this does give a valid analysis in the resource selection context (Manly *et al.*, 1993). Actually, Erickson *et al.* (1998) did something different. They compared the used circles with a systematic sample of 5170 available circles, arguing that this is better than using a random sample of available circles with the same sample size. They completed the analysis using bootstrap estimates for the standard errors of the regression coefficients, and did a sensitivity analysis on the choice of the circle size and the number of circles used in the analysis.

My intention in introducing this example is not to criticise the analysis, which is about the best that is known at present. However, there are some obvious questions about the analysis that are difficult to answer at this time. Here are a few:

In general, with a huge number of possible data points, how do you decide how many are enough to use?

How effective is bootstrapping to calculate standard errors that incorporate uncertainty due to model selection as well as sampling errors?

How should we decide on the 'best' circle size in applications like this?

Does the use of different buffer sizes have biological meaning because animals are selecting for different characteristics at different scales?

What is the effect on the analysis of the initial pixel size?

My contention with examples like this using GIS is that it is possible to collect large sets of data relatively easily but we are struggling to know how best to analyse the data.

Searching for Patterns in Large Data Sets

There is a new area of analysis developing called data mining. This is concerned with searching data sets for unusual associations or differences. There are a number of approaches used including tree-based regression, the use of neural networks, and so on. Commercial software is available for doing the calculations. My impression is that most of the work in this area is being done by information scientists rather than statisticians, and I wonder why this is so.

I suspect that one of the problems is that statisticians tend to be rather conservative in what they regard their subject to be about. I agree with the following statement by Maindonald (1999):

Academic statistics must enlarge its purview to take in all aspects of data collection, data manipulation and data interpretation. Unless this happens other more strongly computer-based disciplines will offer that broader perspective, taking what they need from statistics, and consigning academic statistics to a supporting role.

This leads me on to the third topic that I want to discuss.

4. The Use of Statistics by Scientists in General

In a recent paper read at a meeting of the Royal Statistical Society in the UK, Lindsey (1999) made the following statement:

Statisticians are greatly concerned about the low public esteem for statistics. The discipline is often viewed as difficult and unnecessary, or at best as a necessary evil.

In the discussion of this paper there were 19 contributors. Only two of them questioned this assertion. This suggests to me that, in the UK at least, many statisticians see themselves as having an 'image' problem. In fact, I think this is true in other countries as well.

I would like to mention what is to me another indicator along the same lines. A few years ago I was asked to represent the New Zealand government on the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR), an international organization involving 28 countries, including the USA, concerned, as its name suggests, with the state of the flora and fauna of the Antarctic. I chose to go to meetings of the Statistics Subgroup of the section on Environmental Monitoring and Management.

Much to my surprise, this group concerned with technical matters to do with data collection and analysis consists mainly of biologists. Only the UK (and now New Zealand) seem to think that it is appropriate to send a statistician to the meetings. I wonder what would be thought of a biological subgroup consisting mainly of statisticians!

The biologists in the subgroup are definitely first class, and leaders in biological research in the Antarctic. However, the questions addressed are either purely statistical or a mixture of biological and statistical ideas. At the first meeting that I attended I discovered that the method that had been used by CCAMLR to detect anomalous years was flawed because of confusion between the standard error of a mean and the standard deviation of a single observation. This is what happens when statistics is not done by statisticians.

Of course, the problem of simple mistakes in analyses being made by non-statisticians has been noted by others. Maindonald (1999) puts it quite forcefully:

My reading of published papers persuades me that serious problems with the design of data collection and with data analysis are common. A cursory overview of papers in major international journals may be sufficient to reveal examples of serious statistical misinterpretation ... there is a case for treating all published analyses as preliminary, pending scrutiny by researchers with relevant statistical skills!

My own experience agrees with this comment.

If it is accepted that statistics has an 'image' problem, what can be done about it? There is no simple answer to this question, but I think that there is a communication problem that needs to be overcome to begin with. In terms of research, I think that most papers published in statistics journals are not used by scientists because the scientists cannot understand the methods, and people tend to ignore what they do not understand. In addition, much of what is done is irrelevant to the scientists anyway.

I suppose that many statisticians, and statistical researchers in particular will consider that this does not matter. I think it does because unless statistics is seen as useful and relevant by the scientific community in general, then the long-term outlook is bleak. Better still, statistics needs to be seen useful and relevant by the general non-scientific public. This is about survival of the subject in the next millenium. I emphasize that this is not an argument against the development of new statistical theory, which is clearly essential.

There is at least one statistical journal that has the right philosophy. I understand that the organisers of the annual Kansas State University Conferences on Applied Statistics in Agriculture had an important role in the founding of the *Journal of Agricultural, Biological and Environmental Statistics*, which has the mission of publishing papers that are truly useful. Somehow or other I have ended up as the current Editor of this journal. I hope that in that role I can have some influence in moving the subject of statistics in the right direction.

Summary

In this paper I have given some personal views about some areas that I see as being important for the future of statistics as a subject.

The first area is computer-intensive methods. I have argued that these are wonderful new tools, but that they are sometimes being used without proper checks on whether they work. I have also noted that we can now use complicated models with large numbers of parameters to analyse data, but I wonder whether the outputs always match reality.

The second area is the analysis of the extremely large data sets that are becoming more common and are stretching the capacity of conventional statistical methods. Here I am concerned that new methods for coping with these data sets seem to be coming from information scientist rather than statisticians. This is a danger to the future of statistics as a subject, and I argue that statisticians may need to develop a wider idea of what their subject encompasses.

Finally, I note that many statisticians seem to agree that the subject has low public esteem. Also, many scientists act as if statisticians are irrelevant for what they do, but often make mistakes using statistics. My opinion in this area is that the future of statistics in the next millennium is only assured if the subject is seen as useful and relevant by the scientific community, and preferably by the general public as well.

References

- Clarke, J., Manly, B.F.J., Kerry, K., Gardner, H. Franchi, E., Corsolini, S. and Focardi, S. (1998). Sex differences in Adéle penguin foraging durations. *Polar Biology* 20: 248-258.
- Cushman, J.H., Lawton, J.L. and Manly, B.F.J. (1993). Latitudinal pattern in European ant assemblages: variations in species richness and body size. *Oecologia* 95: 30-37.
- Dwass, M. (1957). Modified randomization tests for non-parametric hypotheses. *Annals of Mathematical Statistics* 28: 181-187.
- Erickson, W.P., McDonald, T.L. and Skinner, R. (1998). Habitat selection using GIS data: a case study. *Journal of Agricultural, Biological and Environmental Statistics* 3: 296-310 (1998).
- Fisher, R.A. (1935). The Design of Experiments. Oliver and Boyd, Edinburgh.
- Fisher, R.A. (1936). The coefficient of racial likeness and the future of craniometry. *Journal of the Royal Anthropological Institute* 66: 57-63.
- Hall, P. and Wilson, S. (1991). Two guidelines for hypothesis testing. *Biometrics* 47: 757-762.
- Hayes, D.L. (1995). *Recovery Monitoring of Pigeon Guillemot Populations in Prince William Sound, Alaska*. Oil Spill Restoration Report (Restoration Project 94173), US Fish and Wildlife Service, Anchorage, Alaska.
- Hilborn, R., Pascual, M., Gerber, L. and Gulland, F. (1997). *The Impact of Incidental Catch on Hooker's Sea Lions Phocarctos Hookeri*. Report submitted to the New Zealand Department of Conservation, Wellington, New Zealand.
- Lindsey, J.K. (1999). Some statistical heresies. *The Statistician* 48: 1-40.
- Maindonald, J.H. (1999). New approaches to using scientific data - statistics, data mining and related technologies in research and research training. Discussion document on the web at http://www.anu.edu.au/graduate/papers/gs98_2.html.
- Manly, B.F.J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman and Hall, London.
- Manly, B.F.J., McDonald, L.L. and Thomas, D.L. (1993). *Resource Selection by Animals*. Chapman and Hall, London.
- Pallisade (1995). *@RISK Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 123*. Palisade Corporation, 31 Decker Road, Newfield, NY 14867.
- Raup, D.M. and Sepkoski, J.J. (1984). Periodicity of extinctions in the geologic past. *Proceedings of the National Academy of Sciences* 81: 801-805.
- Shaw, G.B. (1944). *Everybody's Political What's What*. Constable, London.
- Sherley, G. and Wakelin, M. (1999). Impact of monofluoroacetate ("1080") on forest invertebrates at Ohakune, North Island, New Zealand. *New Zealand Journal of Zoology* (in press).

Table 1 Experimental design for the 9 and 18 day trials to compare invertebrate numbers on placebo and sodium monofluoroacetate (1080) pellets.

(a) 9 Day Trials			
Grid	Days 1 - 3	Days 4 - 6	Days 7 - 9
Control	Placebo pellets	Placebo pellets	Placebo pellets
Treated	Placebo pellets	1080 pellets	Placebo pellets
(b) 18 Day Trials (as above plus)			
Grid	Days 10 - 12	Days 13 - 15	Days 16 - 18
Control	Placebo pellets	Placebo pellets	Placebo pellets
Treated	Placebo pellets	1080 pellets	Placebo pellets

Table 2 Calculation of E_1 , the immediate effect of poison pellets, and E_2 , the residual effect of poison pellets for the first 9 day part of trial 1 (Before is days 1-3, During is days 4-6, and After is for Day 9 only).

	Average Invertebrates per Bait		
	Before	During	After
Control	0.66	0.78	1.89
Treated	1.01	0.39	0.76
Difference	0.35	-0.39	-1.13
Effect	$E_1 = -0.74$		$E_2 = -1.47$

Table 3 Full results for the poison pellet trials in terms of the statistics E_1 (the immediate effect), E_2 (the residual effect), E_3 (the immediate effect from a repeated application), and E_4 (the residual effect from a repeated application). There were 13 trials for 9 days. Six of these trials were immediately followed by another 9 day trial.

Trial	First 9 Day Trial		Repeated 9 Day Trial	
	E_1	E_2	E_3	E_4
1	-0.74	-1.47	-1.85	-0.52
2	-0.38	-0.98	-1.79	-0.68
3	-0.19	-0.10	-0.90	-0.59
4	-0.11	-0.01	-0.66	-0.18
5	0.09	-0.59		
6	-0.51	0.68		
7	-0.12	0.46		
8	-0.15	0.05		
9	-0.20	0.01		
10	-2.13	0.30	-0.03	-0.53
11	-1.98	-0.51	0.12	-0.89
12	-0.66	0.10		
13	-0.79	0.02		
Mean	-0.60	-0.16	-0.85	-0.56